

Notes on Fano Ratio and Portfolio Optimization

Zura Kakushadze^{§1} and Willie Yu^{#2}

Abstract

We discuss – in what is intended to be a pedagogical fashion – generalized “mean-to-risk” ratios for portfolio optimization. The Sharpe ratio is only one example of such generalized “mean-to-risk” ratios. Another example is what we term the Fano ratio (which, unlike the Sharpe ratio, is independent of the time horizon). Thus, for long-only portfolios optimizing the Fano ratio generally results in a more diversified and less skewed portfolio (compared with optimizing the Sharpe ratio). We give an explicit algorithm for such optimization. We also discuss (Fano-ratio-inspired) long-short strategies that outperform those based on optimizing the Sharpe ratio in our backtests.

JEL Classification numbers: G00

Keywords: Portfolio; Stocks; Equities; Optimization; Sharpe Ratio; Fano Ratio; Risk; Return; Expected Return; Alpha; Specific Risk; Idiosyncratic Risk; Factor

¹ Zura Kakushadze, Ph.D., is the President of Quantigic® Solutions LLC, and a Full Professor at Free University of Tbilisi.

[§] Quantigic® Solutions LLC, 1127 High Ridge Road #135, Stamford, CT 06905
Email: zura@quantigic.com

² Willie Yu, Ph.D., is a Research Fellow at Duke-NUS Medical School.
Email: willie.yu@duke-nus.edu.sg

DISCLAIMER: This address is used by the corresponding author for no purpose other than to indicate his professional affiliation as is customary in publications. In particular, the contents of this paper are not intended as an investment, legal, tax or any other such advice, and in no way represent views of Quantigic® Solutions LLC, the website www.quantigic.com or any of their other affiliates.

Loadings; Factor Covariance Matrix; Risk Factor; Volatility; Variance; Covariance; Correlation; Bounds; Trading Costs; Constraints; Regression; Weights

1 Introduction and Summary

When constructing a (e.g., stock) portfolio, one balances risk and reward (i.e., expected return) [36]. Mean-variance optimization [28] provides an implementation of this general idea. In some (somewhat limited) sense, maximizing the Sharpe ratio [36] can be taken as a justification for mean-variance optimization. Thus, without costs, bounds, constraints, etc., maximizing the Sharpe ratio of a portfolio is equivalent to mean-variance optimization. However, once, e.g., costs are included, this equivalence is gone. This begs the question:³

Can we anchor portfolio optimization on quantities other than the Sharpe ratio? In these notes we address precisely this question. The Sharpe ratio is a ratio of the (properly adjusted – see below) expected return over the standard deviation. So, it is a ratio of the expected return to a *particular measure* of risk, in this case, the standard deviation. However, we can consider other measures of risk, e.g., some generic function of the variance.⁴ Thus, one property of the Sharpe ratio is that it depends on the time horizon for which it is calculated. E.g., a daily expected return and volatility give us a daily Sharpe ratio, which is on average lower (by a factor of $\sqrt{252}$, where 252 is the approximate number of trading days in a year, if we focus on stocks) than an annualized Sharpe ratio. If the daily expected return and volatility are stable in time, the Sharpe ratio goes to infinity as \sqrt{T} with the time horizon T .

In contrast, the mean-to-variance ratio (i.e., the expected-return-to-variance ratio) – which we refer to as the Fano ratio (see the next section) – is independent of the horizon T (in the aforementioned sense). We could then take the Fano ratio as the starting point for portfolio optimization. As mentioned above, more generally, we can take a ratio of the expected return to a suitable function of the variance. This is the avenue we explore in these notes, which are intended to be pedagogical.

In Section 2 we discuss maximizing generalized mean-to-risk ratios in the

³ Modifications of mean-variance optimization have been discussed before; see, e.g., [25], [34], [7], [29], [8]. E.g., in [25] the standard deviation is replaced by MAD (that is, mean absolute deviation). Here we take a rather different approach.

⁴ The standard deviation is a square root of the variance, but other functions are also possible.

context of long-only portfolios. Maximizing the Fano ratio leads to simplifications (compared with the general case). Dealing with nonnegativity of the portfolio weights, just as when maximizing the Sharpe ratio, requires an iterative procedure and we provide an approximate relaxation algorithm for optimizing the Fano ratio. For long-only portfolios optimizing the Fano ratio effectively amounts to shifting the expected returns by a positive amount, which results in fewer stocks being excluded from the portfolio (including some stocks with negative expected returns), i.e., in a more diversified and less skewed portfolio (compared with optimizing the Sharpe ratio).

In Section 3 we discuss long-short portfolios, for which certain issues with optimizing the Fano ratio inspire construction of new “multiply-optimized” strategies, which outperform optimizing the Sharpe ratio. We briefly conclude in Section 4.

2 Generalized Mean-to-Risk Ratios

Our discussion below is agnostic to the underlying tradable instruments, which a priori can be stocks, bonds, currencies, etc. However, for the sake of definiteness, let us focus on a portfolio of stocks (e.g., 2,000+ most liquid U.S. stocks). So, we have N stocks with time series of (e.g., close-to-close daily, weekly, monthly or some other horizon) returns R_{is} , $i = 1, \dots, N$. Here the index $s = 1, \dots, T$ labels trading days on which these returns are computed ($s = 1$ labels the most recent date).⁵

Above, R_{is} are the realized returns (ex-post). We can also define expected returns (ex-ante) via, e.g., moving averages:

$$E_{is} = \frac{1}{d} \sum_{s'=s+1}^{s+d} R_{is'} \quad (1)$$

Thus, if R_{is} are daily returns, then E_{is} are d -day moving averages. We emphasize that (1) is only an example and there are myriad other ways of constructing E_{is} . Generally, expected returns can be quite convoluted and have no simple financial interpretation, e.g., machine learning based expected returns [18]. In the following, for the sake of simplicity, we will omit the index s and refer to expected returns as E_i . Thus, we can think of E_i as the expected returns for $s = 1$ (i.e., “today’s” date). What is important is that E_i are computed out-of-sample.

⁵ Here the returns R_{is} are defined as excess returns w.r.t. a risk-free return. In the case of dollar-neutral portfolios this is not crucial. However, here we do not require dollar neutrality.

Next, we can define a sample covariance matrix C_{ij} based on the time series E_{is} or R_{is} and also computed out-of-sample (ex-ante). In what follows it may appear natural to compute C_{ij} based on the expected returns E_{is} as opposed to the realized returns R_{is} . However, in practice, in many cases it can be (much) simpler to compute C_{ij} based on R_{is} . In some cases basing C_{ij} on E_{is} may not even be practicable. One issue is that typically the lookback – i.e., the number of datapoints in the time series, call it M – is insufficient to compute C_{ij} reliably. Thus, if $M < N + 1$, then the sample covariance matrix C_{ij} is singular, whereas for our purposes below C_{ij} must be positive-definite. Furthermore, unless $M \gg N$, which is rarely if ever the case in practice, the off-diagonal elements (in particular, the correlations – the diagonal elements are relatively stable) are highly unstable out-of-sample rendering C_{ij} essentially useless (unpredictive out-of-sample). So, in practice one replaces the sample covariance matrix C_{ij} via a model covariance matrix, call it Γ_{ij} , such as a multifactor risk model.⁶ If built in-house, Γ_{ij} could a priori be built based on E_{is} (among other things). If it is a third-party product, naturally, it is built based on R_{is} (or some other returns). In any event, here we will not delve into how Γ_{ij} is built. We will simply assume that C_{ij} below is identified with some model covariance matrix Γ_{ij} , which is i) positive-definite and ii) sufficiently stable out-of-sample.

2.1 Generalized Mean-to-Risk Ratios for Portfolios

Now we can define portfolio risk. Let us assume that our portfolio consists of our N stocks with weights w_i . A priori some of these weights can be 0 or negative. The normalization condition for the weights is

$$\sum_{i=1}^N |w_i| = 1 \quad (2)$$

Below we will consider a case with nonnegative weights; for now w_i are general.

The expected return of the portfolio is given by

$$E = \sum_{i=1}^N w_i E_i \quad (3)$$

The expected variance of the portfolio is given by

$$V = \sum_{i,j=1}^N C_{ij} w_i w_j \quad (4)$$

⁶ For a general discussion, see, e.g., [13]. For an explicit open-source implementation of a general multifactor risk model for equities, see [20].

We can define the Sharpe ratio [37] of the portfolio via

$$S = \frac{E}{\sqrt{V}} \quad (5)$$

A nice thing about the Sharpe ratio is that it is invariant under the formal rescalings $w_i \rightarrow \zeta w_i$, where $\zeta > 0$. This rescaling invariance is the reason why in the absence of trading costs, bounds, etc., maximizing the Sharpe ratio is equivalent to mean-variance optimization [28].⁷ Indeed, maximizing (5) (i.e., we find the maximum of S w.r.t. w_i)⁸ is equivalent to maximizing (w.r.t. w_i) the objective function

$$g = E - \frac{\lambda}{2} V \quad (6)$$

and λ is fixed (after maximization) via (2). So, in some (somewhat limited) sense, maximizing the Sharpe ratio can be taken as a justification for mean-variance optimization. However, once, e.g., trading costs, etc., are added, maximizing the Sharpe ratio is no longer equivalent to mean-variance optimization [16].

Furthermore, one property of the Sharpe ratio is that it depends on the time horizon for which it is calculated. E.g., a daily expected return and volatility give us a daily Sharpe ratio, which is on average lower (by a factor of \sqrt{d} , where $d \approx 252$ is the approximate number of trading days in a year, if we focus on stocks) than an annualized Sharpe ratio. If the daily expected return and volatility do not change much in time, then the Sharpe ratio goes to infinity as \sqrt{T} with the time horizon T . So, a practical way of thinking about the Sharpe ratio is that, if, say, the annualized Sharpe ratio is 2, then the probability of losing money in a given year is less than about 2.3% (assuming normally distributed realized returns, that is, which can be farfetched – see below).⁹ Can we define a ratio independent of the time horizon?¹⁰

⁷ More precisely, there is a single exception to this, which is the case of linear costs for establishing trades – see [16] for details.

⁸ Which can be done by ignoring (2) due to the aforesaid rescaling invariance as we can always rescale the weights obtained via such maximization to conform to (2). In fact, maximizing (5) fixes w_i only up to an overall normalization factor.

⁹ Recall the 68-95-99.7 rule: if x is a normally distributed variable with mean μ and standard deviation σ , then we have the following probabilities: $\Pr(\mu - n\sigma \leq x \leq \mu + n\sigma) = P_n$, $P_1 \approx 68.27\%$, $P_2 \approx 95.45\%$, $P_3 \approx 99.73\%$. The probability of losing money when the Sharpe ratio equals n (i.e., $\mu = n\sigma$) then is $\tilde{P} = (1 - P_n)/2$. So, we have $\tilde{P}_1 \approx 15.9\%$, $\tilde{P}_2 \approx 2.3\%$, $\tilde{P}_3 \approx 0.14\%$. However, this does not take into account leverage, margin calls, investor withdrawals and other such nuances.

¹⁰ Another known issue with the Sharpe ratio maximization and mean-variance optimization is that one can get portfolios with low degree of diversification. E.g., consider a simple example

The answer is affirmative. We can simply define the mean-to-variance ratio:¹¹

$$F = \frac{E}{V} \quad (7)$$

This ratio is independent of the horizon T (in the aforementioned sense). Other than tautological “mean-to-variance ratio”, this ratio apparently has not been named in finance. It would appear appropriate to term it *the Fano ratio* due to its relationship to the Fano factor [12] named after Ugo Fano, an Italian American physicist. The Fano factor (a.k.a. variance-to-mean ratio and index of dispersion), in our notations, is simply V/E , i.e., the inverse of the Fano ratio (7).¹² Up to a factor of 2, it is the same as the ratio

$$\kappa = \frac{2E}{V} \quad (8)$$

discussed in [19] in the context of stock price bubbles, where it was argued that the dimensionless ratio κ can be used to define a criterion for when a stock (or a similar instrument) is not a good investment in the long term, which can happen even if the expected return is positive. Thus, assuming log-normal distribution for stock prices, this criterion (i.e., that the stock is not a good investment in a long run) is

$$\kappa < 1 \quad (9)$$

This criterion for the Fano ratio (defined for a single stock) would be $F < 1/2$.

So, naturally, we can ask: why not maximize the Fano ratio (instead of the Sharpe ratio)? In fact, we can define more general “mean-to-risk” ratios via

$$G = \frac{E}{f(V)} \quad (10)$$

where $f(V)$ is some function of V . We can then maximize G instead of S (or F).

Recall, however, that we could ignore the normalization condition (2) when maximizing the Sharpe ratio as the latter is invariant under the rescalings $w_i \rightarrow \zeta w_i$. Such invariance is gone in the case of the Fano ratio or more general ratios G

where all $E_i > 0$ and the matrix $C_{ij} = \sigma_i^2 \delta_{ij}$ is diagonal (uncorrelated returns – this is not a crucial assumption here, the following can happen even for correlated returns). The weights that maximize the Sharpe ratio are given by $w_i = \gamma E_i / \sigma_i^2$, where γ is fixed via (2). Now consider a case where all E_i are small except for one. Then we can have all weights but one small and most of the investment will be allocated to the corresponding single stock thereby forgoing diversification.

¹¹ A mean-to-variance ratio test is advocated as a complement to a coefficient of variation test and a Sharpe ratio test in [4]. Also see [3].

¹² Similarly, the Sharpe ratio is the inverse of the coefficient of variation \sqrt{V}/E .

defined via (10). So, the maximization problem becomes more nontrivial. Thus, we must maximize the objective function

$$g = G + \mu \left(\sum_{i=1}^N |w_i| - 1 \right) \quad (11)$$

where μ is a Lagrange multiplier. The modulus complicates things quite a bit (see Section 3). Therefore, for the sake of simplicity, for now let us focus on the case of long-only portfolios, where $w_i \geq 0$. Then our objective function simplifies:

$$g = G + \mu \left(\sum_{i=1}^N w_i - 1 \right) \quad (12)$$

However, now we have bounds

$$w_i \geq 0 \quad (13)$$

To get a flavor of the problem at hand, at first we will ignore the bounds (13) when solving the maximization problem and then incorporate them via a certain trick.

2.2 Maximization Ignoring Bounds

Maximizing (12) w.r.t. w_i and μ (and ignoring the bounds (13)), we get the following solution ($f'(V)$ is the first derivative w.r.t. V , and C_{ij}^{-1} is the inverse of C_{ij}):

$$w_i = a \hat{\eta}_i + b \hat{\nu}_i \quad (14)$$

$$\hat{\eta}_i = \sum_{j=1}^N C_{ij}^{-1} E_j \quad (15)$$

$$\hat{\nu}_i = \sum_{j=1}^N C_{ij}^{-1} \nu_j \quad (16)$$

$$a = \frac{f(V)}{2E f'(V)} \quad (17)$$

$$b = V - \frac{f(V)}{2f'(V)} \quad (18)$$

Here $\nu_i \equiv 1$ is the unit N -vector (using which might appear redundant at first, but will be useful later). So, we have three unknowns, E , V and a . Using the condition $\sum_{i=1}^N w_i = 1$ and the definitions (3) and (4), we have the following

equations:¹³

$$a\gamma + b\beta^2 = 1 \quad (19)$$

$$E = a\alpha^2 + b\gamma \quad (20)$$

$$V = a^2\alpha^2 + 2ab\gamma + b^2\beta^2 = a^2\alpha^2 + (a\gamma + 1)b = a^2\alpha^2 + (1 - a^2\gamma^2)/\beta^2 \quad (21)$$

where

$$\alpha^2 = \sum_{i,j=1}^N C_{ij}^{-1} E_i E_j \quad (22)$$

$$\beta^2 = \sum_{i,j=1}^N C_{ij}^{-1} \nu_i \nu_j \quad (23)$$

$$\gamma = \sum_{i,j=1}^N C_{ij}^{-1} E_i \nu_j \quad (24)$$

and in (21) we repeatedly used (19). So we can express a via V :

$$a^2 = \frac{V\beta^2 - 1}{\alpha^2\beta^2 - \gamma^2} \quad (25)$$

Combining (19), (18) and (25), we get the following equation involving V only:

$$\gamma^2 \frac{V\beta^2 - 1}{\alpha^2\beta^2 - \gamma^2} = \left(1 + \beta^2 \left[\frac{f(V)}{2f'(V)} - V \right] \right)^2 \quad (26)$$

For general $f(V)$ this equation is transcendental. In some cases it simplifies.

Let us start with the case of the Sharpe ratio $f(V) = \sqrt{V}$. Then we have the familiar solution $b = 0$ and $a = 1/\gamma$. If $f(V) = V^p$, where $p > 0$ and $p \neq 1/2$, then (26) is a quadratic equation for V and can be readily solved. It is also a quadratic equation when $f(V) = \exp(\xi V)$. For $f(V) = \exp(\xi V^2)$ the equation is quartic.

2.2.1 Maximizing Fano Ratio

While the aforesaid quadratic equations can be solved, generally they involve radicals and are not particularly illuminating. However, in the case of the Fano ratio, i.e., when $f(V) = V$, things further simplify and there are no radicals. The solution is:

$$a = \frac{1}{\alpha\beta + \gamma} \quad (27)$$

$$b = \frac{\alpha}{\beta(\alpha\beta + \gamma)} \quad (28)$$

¹³ Note that (20) follows from (17), (18), (19) and (21).

and we have

$$E = \frac{\alpha}{\beta} \quad (29)$$

$$V = \frac{2\alpha}{\beta(\alpha\beta + \gamma)} \quad (30)$$

$$F = \frac{1}{2}(\alpha\beta + \gamma) \quad (31)$$

$$S = \sqrt{\frac{\alpha(\alpha\beta + \gamma)}{2\beta}} \quad (32)$$

Had we maximized the Sharpe ratio, we would get $E_* = \alpha^2/\gamma$, $V_* = \alpha^2/\gamma^2$, $F_* = \gamma$ and $S_* = \alpha$. Since $\gamma < \alpha\beta$, it follows that $S < S_*$ and $F > F_*$ (which should come as no surprise as F and S_* are the maximum possible values thereof), and $E < E_*$ and $V < V_*$. So, maximizing the Fano ratio produces a portfolio with a lower expected return but also a lower expected volatility than maximizing the Sharpe ratio.¹⁴

2.3 Incorporating Bounds

The solution (14) is not necessarily good in the sense that some w_i might be negative. Indeed, even if all E_i are nonnegative, we can have negative w_i due to the off-diagonal elements in C_{ij} . So, we must incorporate the bounds (13) into the solution somehow.

The issue is that we are not dealing with a quadratic optimization problem here.¹⁵ However, not all is lost and the following trick provides a reasonable approximation. Thus, the solution (14) *formally* can be thought of as the solution to maximizing the following quadratic objective function (λ is fixed after solving for w_i by rescaling them such that $\sum_{i=1}^N w_i = 1$, which rescaling is not affected by the bounds (13))

$$\tilde{g} = \tilde{E} - \frac{\lambda}{2} V \quad (33)$$

¹⁴ Note that in portfolios based on maximizing the Sharpe ratio the realized expected return and Sharpe ratio can be vastly different from their expected values based on optimization. Therefore, the fact that the expected return is higher when we maximize the Sharpe ratio as compared to when we maximize the Fano ratio means little in terms of what the realized return will be.

¹⁵ This is also the case when maximizing the Sharpe ratio in the presence of bounds.

subject to the bounds (13), where (note that $\lambda = \alpha(J)\beta(J) + \gamma(J)$)

$$\tilde{E} = \sum_{i=1}^N w_i \tilde{E}_i \quad (34)$$

$$\tilde{E}_i = E_i + E(J) \nu_i \quad (35)$$

$$E(J) = \frac{\alpha(J)}{\beta(J)} \quad (36)$$

$$[\alpha(J)]^2 = \sum_{i,j \in J} [C(J)]_{ij}^{-1} E_i E_j \quad (37)$$

$$[\beta(J)]^2 = \sum_{i,j \in J} [C(J)]_{ij}^{-1} \nu_i \nu_j \quad (38)$$

$$\gamma(J) = \sum_{i,j \in J} [C(J)]_{ij}^{-1} E_i \nu_j \quad (39)$$

where $J = \{i | w_i > 0\}$ is the subset of positive weights, $[C(J)]_{ij}^{-1}$ is the $N(J) \times N(J)$ matrix inverse to the $N(J) \times N(J)$ matrix $[C(J)]_{ij}$ obtained from C_{ij} by restricting $i, j \in J$. (Here $N(J) = |J|$ is the number of elements in J .) So, the catch is that $E(J)$, and thereby \tilde{E}_i , depend on J , which is unknown. Had $E(J)$ been known a priori, then we would simply minimize (33) subject to the bounds (13) via standard quadratic optimization techniques.¹⁶ So, here is a relaxation algorithm that *approximates* the optimal solution. At the initial iteration, we assume that $J^{(0)}$ is the full set $\{1, \dots, N\}$ and compute w_i via (14). If all $w_i \geq 0$, then there is nothing else to do, we are done. So, let us assume that the set $\tilde{J}^{(0)} = \{i | w_i < 0\}$ is not empty. Let us take the value of $\ell \in \tilde{J}^{(0)}$ for which $F_\ell = \min(F_i)$, where $F_i = E_i/C_{ii}$ are the Fano ratios for each stock.¹⁷ We then permanently set $w_\ell = 0$, take $J^{(1)} = J^{(0)} \setminus \ell$ and compute w_i via (14). If the resulting $w_i \geq 0$ for all $i \in J^{(1)}$, then we are done. So, let us assume that the set $\tilde{J}^{(1)} = \{i | w_i < 0, i \in J^{(1)}\}$ is not empty. Let us take the value of $\ell' \in \tilde{J}^{(1)}$ for which $F_{\ell'} = \min(F_i)$, $i \in J^{(1)}$ (see fn.17). We then permanently set $w_{\ell'} = 0$, take $J^{(2)} = J^{(1)} \setminus \ell'$ and compute w_i via (14). And so on. We repeat this procedure until at some k -th iteration all $w_i \geq 0$ for $i \in J^{(k)}$. As always, one issue with this relaxation algorithm is the computational cost: we must compute the inverse matrix $[C(J)]_{ij}^{-1}$ at each iteration. However, for a K -factor model of the form

¹⁶ See, e.g., [10], [11], [15], [16], [31], [32], and references therein.

¹⁷ If there are multiple values of ℓ for which $F_\ell = \min(F_i)$, then we take the value of ℓ for which $C_{\ell\ell} = \max(C_{ii})$, and if there are still multiple values of ℓ remaining, we simply take the lowest ℓ .

(here ξ_i^2 is the specific a.k.a. idiosyncratic risk, Ω_{iA} , $A = 1, \dots, K$ is the factor loadings matrix, and ϕ_{AB} is the factor covariance matrix)

$$C_{ij} = \xi_i^2 \delta_{ij} + \sum_{A=1}^K \Omega_{iA} \phi_{AB} \Omega_{jA} \quad (40)$$

to compute $[C(J)]_{ij}^{-1}$, we only need to invert the $K \times K$ matrix ϕ_{AB} once, plus we must invert a $K \times K$ matrix¹⁸

$$[Q(J)]_{AB} = \phi_{AB}^{-1} + \sum_{i \in J} \frac{1}{\xi_i^2} \Omega_{iA} \Omega_{iB} \quad (41)$$

at each iteration. However, these inversions are much cheaper assuming $K \ll N$.

2.4 The “Market” Mode

The issue we wish to address next is equally pertinent to maximizing the Sharpe ratio, the Fano ratio and the generalized mean-to-risk ratios we discuss above. For the sake of definiteness and simplicity, let us focus on the case of maximizing the Sharpe ratio. Let us ignore the bounds (13) for a moment. Then the weights are given by

$$w_i = a \sum_{j=1}^N C_{ij}^{-1} E_j \quad (42)$$

$$a^{-1} = \sum_{i,j=1}^N C_{ij}^{-1} E_i \nu_j \quad (43)$$

For a typical configuration, even if all E_i are nonnegative, close to 50% of the weights w_i can be negative. Thus, to illustrate this point, let us consider the following “toy” covariance matrix:¹⁹ $C_{ij} = \sigma_i \sigma_j \Psi_{ij}$, where σ_i^2 are the variances; the correlation matrix $\Psi_{ij} = (1 - \rho) \delta_{ij} + \rho \nu_i \nu_j$; and $\nu_i \equiv 1$ is the unit N -vector. I.e., all N stocks have uniform pair-wise correlations equal ρ . Inverting this matrix gives the following weights:

$$w_i = \frac{a}{\sigma_i (1 - \rho)} \left[\tilde{E}_i - \frac{\rho \nu_i}{1 + (N - 1)\rho} \sum_{j=1}^N \tilde{E}_j \right] \quad (44)$$

¹⁸ Iteratively dropping the stock with the lowest Fano ratio is only an approximation. However, it is computationally feasible. E.g., iteratively dropping the stock with the smallest impact on the full Fano ratio (31) would prohibitively require inverting $\sim N$ matrices (41) at each iteration.

¹⁹ This is an example of a 1-factor model.

where $\tilde{E}_i = E_i/\sigma_i$ are the normalized expected returns. Generically, the latter are expected to be roughly symmetrically distributed around their mean. It is then evident from (44) that, unless $\rho \ll N$, roughly 50% of the weights w_i are negative.

Now, in practice the correlation matrix with uniform pair-wise correlations is unrealistic. However, the above issue persists even for realistic correlation matrices. Thus, consider a general correlation matrix Ψ_{ij} . We can always write it as

$$\Psi_{ij} = (1 - \rho) \delta_{ij} + \rho \nu_i \nu_j + \Delta_{ij} = \Psi' + \Delta_{ij} \quad (45)$$

Here $\rho = \frac{1}{N(N-1)} \sum_{i,j=1; i \neq j}^N \Psi_{ij}$ is the *average* pair-wise correlation, and $\sum_{i,j=1}^N \Delta_{ij} = 0$.

In the zeroth approximation we can drop Δ_{ij} , i.e., $\Psi_{ij} \approx \Psi'_{ij}$. Its first principal component $U_i^{(1)} = \nu_i/\sqrt{N}$. It describes the “market” mode [6],²⁰ i.e., the average correlation of all stocks, which is nonzero (and not small, definitely $\rho \not\ll 1/N$).²¹ The “market” mode corresponds to the overall movement of the broad market, which affects all stocks (to varying degrees) – cash inflow (outflow) into (from) the market tends to push stock prices higher (lower). This is the market risk factor. To mitigate this risk factor, one can, e.g., hold a dollar-neutral portfolio of stocks. However, long-only portfolios are exposed to market risk by construction.

So, neutralizing the market risk factor while maximizing the Sharpe ratio is an unwelcome feature. Why? Because at the end we impose the bounds (13) anyway, so the market risk is still present, but the resultant portfolio gets artificially distorted due to pushing the negative weights (that is, in the unbounded optimization) to zero thereby also affecting the positive weights. The culprit here is that, when maximizing the Sharpe ratio using a covariance matrix that includes the “market” mode, we approximately neutralize the portfolio w.r.t. the market risk. Put differently, we hedge against all stocks – i.e., the broad market – going bust. However, holding a long-only portfolio with thousands of stocks invariably is exposed to the broad market. So, we must eliminate the “market” mode out of the covariance matrix.²²

²⁰ Also see [22].

²¹ Note that the eigenvalue of Ψ'_{ij} corresponding to $U_i^{(1)}$ is $\lambda'_* = 1 + \rho(N-1)$.

²² Here one can argue that one can build a “market-neutral” long-only portfolio by picking stock weights such that they are neutral w.r.t. market betas by utilizing the fact that some betas can be negative. However, not only do the betas tend to be highly unstable out-of-sample, we still have the bounds (13) (which are not that easy to satisfy for beta-neutral portfolios), and neutralizing against the “market” mode (whose elements are all positive) in no way is helpful in building a beta-neutral long-only portfolio. For a review of some market-neutral strategies (which, however, are not long-only), see, e.g., [26] and references therein.

In the context of factor models (40) this can be achieved relatively easily. In this case (ignoring the bounds (13)) we have (as above, a is fixed via $\sum_{i=1}^N w_i = 1$)

$$w_i = a \left[\frac{E_i}{\xi_i^2} - \frac{1}{\xi_i^2} \sum_{A,B=1}^K \Omega_{iA} Q_{AB}^{-1} \sum_{j=1}^N \frac{E_j}{\xi_j^2} \Omega_{jB} \right] \quad (46)$$

$$Q_{AB} = \phi_{AB}^{-1} + \sum_{i=1}^N \frac{1}{\xi_i^2} \Omega_{iA} \Omega_{iB} \quad (47)$$

Eliminating the “market” mode from C_{ij} then amounts to requiring that the factor loadings matrix is orthogonal to some *positive* N -vector $v_i > 0$:

$$\sum_{i=1}^N v_i \Omega_{iA} \equiv 0 \quad (48)$$

Then we no longer have roughly 50% of negative weights w_i . While some of these weights can still be negative, typically the number of such negative weights will be relatively small compared with N (assuming all $E_i \geq 0$, that is). So, what are v_i ?

One – but not the only – way of thinking about v_i is that they are the weights of some *benchmark* long-only portfolio: $v_i = w_i^{\text{benchmark}} > 0$. The choice of this benchmark portfolio is not that critical provided it is reasonably diversified. For example, we can take²³ $v_i \equiv 1$, i.e., an equally-weighted benchmark. We can take $v_i = 1/\sigma_i^2$ or $v_i = 1/\xi_i^2$. This can skew the portfolio toward low-volatility (which are typically large market cap) stocks. To mitigate this, we can Windsorize or otherwise deal with the tails in the skewed (roughly log-normal) distribution of σ_i (or ξ_i). Etc.

2.5 Statistical Risk Models

Statistical risk models [23] provide a particularly simple example of factor models, where the factor covariance matrix is diagonal. Thus, let $\Psi_{ij}^{\text{sample}}$ be the sample correlation matrix computed based on time series of historical returns. $\Psi_{ij}^{\text{sample}}$ can be singular. This will not affect our discussion below. The sample covariance matrix is $C_{ij}^{\text{sample}} = \sigma_i \sigma_j \Psi_{ij}^{\text{sample}}$. We can construct a statistical risk model

²³ Up to an overall normalization, that is.

covariance matrix C_{ij} as follows:

$$C_{ij} = \sigma_i \sigma_j \Psi_{ij} \quad (49)$$

$$\Psi_{ij} = \tilde{\xi}_i^2 \delta_{ij} + \sum_{a=1}^K \lambda^{(a)} V_i^{(a)} V_j^{(a)} \quad (50)$$

$$\tilde{\xi}_i^2 = 1 - \sum_{a=K+1}^r \lambda^{(a)} [V_i^{(a)}]^2 \quad (51)$$

Here $V_i^{(a)}$ are the principal components of the matrix Ψ_{ij}^{sample} with the corresponding eigenvalues in the descending order: $\lambda^{(1)} > \lambda^{(2)} > \dots > \lambda^{(r)}$, where r is the rank of Ψ_{ij}^{sample} (if $r < N$, then for $a > r$ we have $\lambda^{(a)} = 0$). The number of factors K is determined via (truncated or rounded) eRank (effective rank) [35] – see [23] for details. The issue with the so-constructed C_{ij} is that it contains the “market” mode. Indeed, without loss of generality we can assume that all elements of the first principal component $V_i^{(1)} > 0$ – this can be ensured by, if need be, changing the basis as follows: $C_{ij} \rightarrow \epsilon_i \epsilon_j C_{ij}$, where $\epsilon_i = \text{sign}(V_i^{(1)})$. Then the all-positive $V_i^{(1)}$ can be regarded as the “market” mode [6]. In fact, for large N we have $V_i^{(1)} \approx 1/\sqrt{N}$. Note that higher principal components $V_i^{(a>1)}$ invariably have negative elements. So, we need to eliminate the first principal component. This can be achieved simply by defining

$$C_{ij} = \sigma_i \sigma_j \hat{\Psi}_{ij} \quad (52)$$

$$\hat{\Psi}_{ij} = \hat{\xi}_i^2 \delta_{ij} + \sum_{a=2}^K \lambda^{(a)} V_i^{(a)} V_j^{(a)} \quad (53)$$

$$\hat{\xi}_i^2 = 1 - \lambda^{(1)} [V_i^{(1)}]^2 - \sum_{a=K+1}^r \lambda^{(a)} [V_i^{(a)}]^2 \quad (54)$$

This is not the only possible definition, but it is as good as any other. With this definition we can think of the benchmark portfolio as that with $v_i = V_i^{(1)}/\sigma_i$.

Our discussion above is for maximizing the Sharpe ratio but equally applies to maximizing the Fano ratio and the generalized mean-to-risk ratios. This is because in all these cases the weights involve inverting the covariance matrix. Indeed, in (14) we have $w_i = a \sum_{j=1}^N C_{ij}^{-1} E'_j$, where $E'_i = E_i + \nu_i b/a$, so the above still applies.

2.6 Why Is This Useful?

For long-only portfolios optimizing the Fano ratio *effectively* amounts to shifting the expected returns by a positive amount via (35), which results in fewer stocks

being excluded from the portfolio due to the bounds (13) (including some stocks with negative expected returns E_i , for which the effective returns \tilde{E}_i can be positive), i.e., in a more diversified and less skewed portfolio (compared with optimizing the Sharpe ratio). This is evident for a diagonal matrix $C_{ij} = \sigma_i^2 \delta_{ij}$. And this conclusion persists for non-diagonal C_{ij} of the factor model form with the “market” mode removed. This can be illustrated using a simple 1-factor model of the form $C_{ij} = \sigma_i \sigma_j \Psi_{ij}$, where the correlation matrix $\Psi_{ij} = (1 - \rho) \delta_{ij} + \rho s_i s_j$, and $s_i = 1$ for half of the values of i , and $s_i = -1$ for the other half (the number of stocks N is assumed to be even). (As above, let us ignore the bounds for a moment.) Then we have (here a is given by (27))

$$w_i = \frac{a}{\sigma_i (1 - \rho)} \left[\tilde{E}'_i - \frac{\rho s_i}{1 + (N - 1)\rho} \sum_{j=1}^N \tilde{E}'_j s_j \right] \quad (55)$$

$$\tilde{E}'_i = \tilde{E}_i + \frac{\alpha}{\beta} \frac{\nu_i}{\sigma_i} \quad (56)$$

$$\tilde{E}_i = E_i / \sigma_i \quad (57)$$

$$\alpha^2 = \frac{1}{(1 - \rho)} \left[\sum_{i=1}^N \tilde{E}_i^2 - \frac{\rho}{1 + (N - 1)\rho} \left(\sum_{i=1}^N \tilde{E}_i s_i \right)^2 \right] \quad (58)$$

$$\beta^2 = \frac{1}{(1 - \rho)} \left[\sum_{i=1}^N \frac{1}{\sigma_i^2} - \frac{\rho}{1 + (N - 1)\rho} \left(\sum_{i=1}^N \frac{s_i}{\sigma_i} \right)^2 \right] \quad (59)$$

It is reasonable to assume that there is no substantial correlation between the values of σ_i and the signs s_i , or the values of E_i and s_i . Then we can estimate that $\left| \sum_{i=1}^N s_i / \sigma_i \right| \lesssim \sqrt{N} / \sigma_*$, where $N / \sigma_*^2 = \sum_{i=1}^N 1 / \sigma_i^2$. Similarly, $\left| \sum_{i=1}^N \tilde{E}_i s_i \right| \lesssim \sqrt{N} \tilde{E}_*$, where $N \tilde{E}_*^2 = \sum_{i=1}^N \tilde{E}_i^2$. Further, we can reasonably assume that $\rho \ll 1/N$ (and $N \gg 1$). Then we have

$$\alpha^2 = \frac{N \tilde{E}_*^2}{(1 - \rho)} [1 - \mathcal{O}(1/N)] \quad (60)$$

$$\beta^2 = \frac{N}{\sigma_*^2 (1 - \rho)} [1 - \mathcal{O}(1/N)] \quad (61)$$

and, up to terms suppressed by $1/N$, the weights are given by

$$w_i \approx \frac{a}{\sigma_i (1 - \rho)} \left[\tilde{E}_i + \tilde{E}_* \nu_i \frac{\sigma_*}{\sigma_i} - \frac{s_i}{N} \sum_{j=1}^N \left(\tilde{E}_j s_j + \tilde{E}_* \frac{\sigma_* s_j}{\sigma_j} \right) \right] \quad (62)$$

In this expression the terms containing \tilde{E}_* are pertinent to optimizing the Fano ratio; the other two terms are present when optimizing the Sharpe ratio (in which

case the overall normalization coefficient a is different). And it is precisely the second term in the square brackets in (62) that makes the difference here. Here is why and how.

Based on our argument above, the term in (62) containing a sum over j has a magnitude of order \tilde{E}_*/\sqrt{N} . Now consider the values of the index i such that $\sigma_i \ll \sqrt{N} \sigma_*$. For such i the second term in (62) dominates the term containing the sum and we have

$$w_i \approx \frac{a}{\sigma_i(1-\rho)} \left[\tilde{E}_i + \tilde{E}_* \nu_i \frac{\sigma_*}{\sigma_i} \right] \quad (63)$$

For such values of i , w_i can be positive even for negative returns E_i . This is because i) the contributions of the off-diagonal terms in the covariance matrix C_{ij} into the optimization are suppressed for such i , and ii) the intrinsic-to-Fano-ratio term (proportional to \tilde{E}_*) provides an additive positive contribution. This reduces the number of stocks with negative weights (when we ignore the bounds, that is), which are then “pushed up” when we include the bounds. And this additive contribution is positive even for the values of i for which $\sigma_i \ll \sqrt{N} \sigma_*$. Let us quantify this.

We can reasonably assume that there is no substantial correlation between \tilde{E}_i and $1/\sigma_i$. Then the deviations for the two terms in the parenthesis in (62) can be estimated independently. The standard deviation of the term containing the sum in (62) (approximately) is $\sqrt{2/N} \tilde{E}_*$. Conservatively, assuming that its actual value deviates by $5/\sqrt{2} \approx 3.54$ standard deviations in either direction, we can estimate the bound $\tilde{\sigma}$ on σ_i such that, for $\sigma_i < \tilde{\sigma}$ it is unlikely (with roughly 3.54 standard deviations confidence level) that the term containing the sum in (62) outweighs the second term in (62) such that the total contribution of these terms is negative:

$$\tilde{\sigma} \approx \frac{\sqrt{N}}{5} \sigma_* \quad (64)$$

And the number of such stocks typically is pretty small (compared with the number of stocks in the portfolio). To illustrate this, here is an example from data. We take the data for the universe of tickers as of Sep 6, 2014 that have historical pricing data on <http://finance.yahoo.com> (accessed on Sep 6, 2014) for the period Aug 1, 2008 through Sep 5, 2014.²⁴ We restrict this universe to include only U.S. listed common stocks and class shares (no OTCs, preferred shares, etc.) with BICS (Bloomberg Industry Classification System) sector, industry and sub-industry assignments as of Sep 6, 2014. The number of such tickers in our data is

²⁴ The choice of this window is not critical here. We simply used data readily available to us.

3811. We then compute the 21-trading-day (i.e., 1-month) historical volatilities based on daily close-to-close returns for the most recent date in the data, Sep 5, 2014. One stock was not trading (zero volatility) in that 21-trading day period, so we are left with 3810 stocks with nonzero volatilities. These are our σ_i . The cross-sectional distribution of σ_i is roughly log-normal, with a long tail at higher values (see Figures 1 and 2). The summary of these 3810 values of σ_i is as follows: Min = 0.64×10^{-3} , 1st Quartile = 9.16×10^{-3} , Median = 0.0137, Mean = 0.0185, 3rd Quartile = 0.02197, Max = 0.3252, SD (standard deviation) = 0.01706, MAD (mean absolute deviation) = 8.37×10^{-3} . Further, we have $\sigma_* = 9.99 \times 10^{-3}$, and the number of stocks in this universe with $\sigma_i \geq \tilde{\sigma}$ is only 14. If we take 10 in the denominator²⁵ instead of 5 in the definition (64), we still only get 78 stocks with $\sigma_i \geq \tilde{\sigma}$. If we restrict our stock universe to the top 2000 most liquid stocks by ADDV (average daily dollar volume, also computed based on the same 21-trading-day period), the results are similar (also see Figures 3 and 4): Min = 0.64×10^{-3} , 1st Quartile = 7.86×10^{-3} , Median = 0.0111, Mean = 0.01470, 3rd Quartile = 0.01678, Max = 0.2566, SD = 0.01408, MAD = 5.60×10^{-3} . Further, we have $\sigma_* = 8.55 \times 10^{-3}$, and the number of stocks in this universe with $\sigma_i \geq \tilde{\sigma}$ is only 16. If we take 10 in the denominator instead of 5 in the definition of (64), again we still only get 69 stocks with $\sigma_i \geq \tilde{\sigma}$.

2.7 Multifactor Risk Models

In the preceding subsection we discuss a simple 1-factor model where the pairwise correlations (after removing the “market” mode) take two values, $\pm\rho$. (If we add back the “market” mode with a uniform correlation ρ_0 , then the pairwise correlations in the resultant correlation matrix take two values, neither of which need be (but one of them can be) negative. Our discussion above can be generalized to multifactor models (with the “market” mode removed). The math is more involved but the gist of it is captured by the 1-factor example we discuss above. Thus, we can reasonably assume that the returns E_i are not significantly correlated with σ_i or the factor loadings Ω_{iA} , so that in optimizing the Fano ratio (as compared with the Sharpe ratio) the expected returns effectively get shifted by a positive additive contribution for most stocks, excepting large volatility stocks. As above, this results in fewer weights violating the bounds (13) and the portfolio is also more diversified.

²⁵ This corresponds to $10/\sqrt{2} \approx 7.07$ standard deviations (instead of $5/\sqrt{2} \approx 3.54$ – see above).

3 Long-Short Portfolios

Above we discuss long-only portfolios. What about long-short portfolios? To maximize the Fano ratio, we need to maximize the objective function (11). The modulus in (11) complicates things. First, its derivative is well-defined for $w_i \neq 0$ and for the subset $J = \{i | w_i \neq 0\}$ the maximization of g in (11) is equivalent to $\partial g / \partial w_i = 0$, $i \in J$. For the sake of simplicity,²⁶ let us assume that all $w_i \neq 0$. Then we have all the same formulas as above for the long-only portfolio (without any bounds as w_i need no longer be nonnegative) except that ν_i is replaced by $\chi_i = \text{sign}(w_i)$. So the analog of (14) now must be solved iteratively. However, here we will not delve into solving this problem (or its subtleties) as there is a more prosaic issue to address.

Ignoring the aforesaid subtleties, the equation we would need to solve iteratively reads (see (14) and the subsequent equations for definitions of a and b)

$$w_i = \sum_{j=1}^N C_{ij}^{-1} [a E_j + b \chi_j] \quad (65)$$

where a and b also depend on χ_i . However, it is not this dependence that is problematic. Instead, it is the presence of signs, i.e., χ_i , in (65). Signs are highly unstable (they “flip-flop” a lot, especially for shorter horizons). To illustrate this, let us simplify things and consider the case of a diagonal covariance matrix $C_{ij} = \sigma_i^2 \delta_{ij}$. Then $\hat{\eta}_i = E_i / \sigma_i^2$ and we can set $\chi_i = \text{sign}(E_i)$. So, for a small E_i (e.g., compared with its historical standard deviation or some suitable multiple thereof), if its sign flips (but the absolute value remains small), we can have a 100% opposite contribution from χ_i into (65). This is the root-cause of the aforesaid instability, which also persists even for non-diagonal C_{ij} (in which case things are simply messier). We can think about this as follows. The weights (65) effectively are the same as linearly combining two strategies. One is based on optimizing the Sharpe ratio for the expected returns E_i . The other is based on optimizing the Sharpe ratio for *binary*²⁷ expected returns $\chi_i = \pm 1$. It should come as no surprise to quant traders that the second strategy is suboptimal.

²⁶ Here we will not delve into the $w_i = 0$ (and other important) subtleties. Such subtleties arise, e.g., in the case of mean-variance optimization with linear costs. For a recent discussion, see, e.g., [16]. For a partial list of related literature, see, e.g., [1], [5], [9], [14], [24], [27], [30], [33], [38], and references therein.

²⁷ For the sake of simplicity, assuming, as above, that all $w_i \neq 0$, that is. If some $w_i = 0$, then the corresponding returns are not binary but trinary (with at most a small number of null returns). However, this does not alter the above conclusion relating to the instability of the signs χ_i .

E.g., if we take $\text{sign}(E_i)$ instead of E_i as binary expected returns, this strategy underperforms the strategy based on optimizing E_i . This is because forecasting just the direction and not the magnitude of the expected returns provides only partial information. Linearly combining such a suboptimal strategy with the strategy based on optimizing the Sharpe ratio then also is suboptimal.

Can we fix this? We can smooth out the sign in $\chi_i = \text{sign}(w_i)$. One way to do this is to replace it by, e.g., a hyperbolic tangent: $\chi_i = \tanh(w_i/\Delta_i)$, where Δ_i are some parameters. In the limit $\Delta_i \rightarrow 0$ we recover $\chi_i = \text{sign}(w_i)$. Introducing N new parameters Δ_i can be unappealing as they can easily turn out to be out-of-sample unstable. We can mitigate this, at least to a degree, by taking uniform $\Delta_i \equiv \Delta$ (however, we will relax this below). We then have

$$w_i = \sum_{j=1}^N C_{ij}^{-1} [a E_j + b \tanh(w_j/\Delta)] \quad (66)$$

We can solve this equation, e.g., by linearizing the hyperbolic tangent, which *formally* amounts to the limit where $\Delta \rightarrow \infty$, $b \rightarrow \infty$, and $\tilde{b} = b/\Delta$ is kept finite:

$$w_i = \sum_{j=1}^N C_{ij}^{-1} [a E_j + \tilde{b} w_j] \quad (67)$$

A formal solution²⁸ reads

$$w_i = a \sum_{j=1}^N \tilde{C}_{ij}^{-1} E_j \quad (68)$$

$$\tilde{C}_{ij} = C_{ij} - \tilde{b} \delta_{ij} \quad (69)$$

The overall normalization parameter a is fixed by requiring the normalization condition (2). However, the parameter \tilde{b} is a priori undetermined. Since we have departed from optimizing the Fano ratio, it is no longer evident what \tilde{b} should be. Instead of trying to fix it “theoretically”, we can take a pragmatic approach and treat \tilde{b} as a free parameter. For $\tilde{b} = 0$ we are simply optimizing the Sharpe ratio. For $\tilde{b} \neq 0$, we are optimizing the Sharpe ratio but with a modified covariance matrix \tilde{C}_{ij} , whose off-diagonal elements are the same as those of C_{ij} , but the diagonal elements (variances) are shifted: they can be increased ($\tilde{b} < 0$) or decreased ($\tilde{b} > 0$).

In this regard, it is instructive to consider the case of nonuniform Δ_i . In this case we still have (68), where now

$$\tilde{C}_{ij} = C_{ij} - \tilde{b}_i \delta_{ij} \quad (70)$$

²⁸ We, yet again, use the adjective “formal” as a and \tilde{b} a priori are undetermined (see below).

and $\tilde{b}_i = b/\Delta_i$. Let us consider a factor model of the form (40). If we set $\tilde{b}_i = \theta \xi_i^2$, where ξ_i^2 are the specific variances and θ is a parameter, then for $\theta = 1$ the matrix \tilde{C}_{ij} is singular. We can invert it in the $\theta \uparrow 1$ limit (in this limit the normalization a goes to 0 such that w_i are actually finite) and the result is that, up to an overall normalization factor (fixed via (2)), the weights w_i are given by ϵ_i/ξ_i^2 , where ϵ_i are the residuals of a cross-sectional regression of E_i over the factor loadings Ω_{iA} with the regression weights $z_i = 1/\xi_i^2$ and no intercept²⁹ [16]. Equivalently, $\tilde{\epsilon}_i = \epsilon_i/\xi_i$ are the residuals of a cross-sectional regression of E_i/ξ_i over the matrix Ω_{iA}/ξ_i with unit regression weights (and no intercept – see above). So, here we are interpolating between optimizing the Sharpe ratio and a (weighted) regression.

Formally, we can view (68) as an infinite series (here $a_p = a \tilde{b}^{p-1}$):

$$w_i = \sum_{p=1}^{\infty} a_p \sum_{j=1}^N C_{ij}^{-1} E_j^{(p)} \quad (71)$$

$$E_i^{(p+1)} = \sum_{j=1}^N C_{ij}^{-1} E_j^{(p)} \quad (72)$$

$$E_i^{(1)} = E_i \quad (73)$$

I.e., this is a combination of “once-optimized”, “twice-optimized”, “trice-optimized”, . . . , strategies. In fact, we can simply forget about how we got this result (which was in an ad hoc and handwaving fashion – however, see below) and take a truncated series

$$w_i = \sum_{p=1}^{n_{opt}} a_p \sum_{j=1}^N C_{ij}^{-1} E_j^{(p)} = a \sum_{j=1}^N C_{ij}^{-1} \hat{E}_j \quad (74)$$

where $\hat{E}_i = \sum_{p=1}^{n_{opt}} \tilde{b}^{p-1} E_i^{(p)}$. Only one of the n_{opt} coefficients a_p is fixed by the normalization condition (2), i.e., we can fix $a_1 = a$. As mentioned above, a priori there is no guiding principle for fixing the \tilde{b} parameter.³⁰ However, we can require that the coefficients a_p have the proper scaling properties under $E_i \rightarrow \zeta E_i$ and $C_{ij} \rightarrow \lambda C_{ij}$, where $\zeta > 0$ and $\lambda > 0$ (so that w_i are invariant under such rescalings):

$$a_p \rightarrow \zeta^{-1} a_p \quad (75)$$

$$a_p \rightarrow \lambda^p a_p \quad (76)$$

²⁹ Unless the intercept is already subsumed in the factor loadings matrix Ω_{iA} , that is.

³⁰ More generally, we can depart from $a_p = a \tilde{b}^{p-1}$ and treat the coefficients a_p as independent. Then we can datamine $a_{p>1}$ and see if they are stable out-of-sample. We will not do this here.

This implies that \tilde{b} is invariant under $E_i \rightarrow \zeta E_i$, and we have $\tilde{b} \rightarrow \lambda \tilde{b}$ under $C_{ij} \rightarrow \lambda C_{ij}$. Consider \tilde{b} of the following form:

$$\tilde{b} = \hat{b} h = \hat{b} \sqrt{\frac{\sum_{i,j=1}^N C_{ij}^{-1} E_i E_j}{\sum_{i,j=1}^N C_{ij}^{-1} E_i^{(2)} E_j^{(2)}}} \quad (77)$$

Then \hat{b} is invariant under both the ζ and λ rescalings. We can therefore treat \hat{b} as a purely numerical coefficient. For instance, for $n_{opt} = 2$ we have

$$w_i = a \left[\sum_{j=1}^N C_{ij}^{-1} E_j + \hat{b} h \sum_{j=1}^N C_{ij}^{-1} E_j^{(2)} \right] \quad (78)$$

I.e., we are combining the “once-optimized” and “twice-optimized” strategies with the relative coefficient controlled by \hat{b} . We discuss a backtest of this strategy below.

3.1 Bells and Whistles

While our w_i in (74) are roughly dollar-neutral (due to the presence of the “market” mode in C_{ij} , which a priori need not be removed for long-short portfolios), they are not exactly dollar-neutral. We may wish our long-short portfolio to be exactly dollar-neutral (e.g., due to risk management/compliance requirements, etc.):

$$\sum_{i=1}^N w_i = 0 \quad (79)$$

More generally, we may wish to impose more than one linear homogeneous constraints

$$\sum_{i=1}^N G_{i\alpha} w_i = 0, \quad \alpha = 1, \dots, m \quad (80)$$

where the columns of the $N \times m$ matrix $G_{i\alpha}$ are linearly independent. Such constraints are readily incorporated in the optimization problem by “padding” the factor loadings matrix with the extra m columns: $\tilde{\Omega}_{i\tilde{A}} = (\Omega_{iA}, G_{i\alpha})$, where the index $\tilde{A} = (A, \alpha) \in H$ now takes $\tilde{K} = |H| = K + m$ values ($H = \{\tilde{A}\}$). We

then have (see, e.g., [16])

$$C_{ij}^{-1} = \frac{1}{\xi_i^2} \delta_{ij} - \sum_{\tilde{A}, \tilde{B} \in H} \frac{\tilde{\Omega}_{i\tilde{A}}}{\xi_i^2} \tilde{Q}_{\tilde{A}\tilde{B}}^{-1} \frac{\tilde{\Omega}_{j\tilde{B}}}{\xi_j^2} \quad (81)$$

$$\tilde{Q}_{\tilde{A}\tilde{B}} = \varphi_{\tilde{A}\tilde{B}} + \sum_{i \in J} \frac{1}{\xi_i^2} \tilde{\Omega}_{i\tilde{A}} \tilde{\Omega}_{i\tilde{B}} \quad (82)$$

$$\varphi_{AB} = \phi_{AB}^{-1}, \quad A, B = 1, \dots, K \quad (83)$$

$$\varphi_{A\alpha} = 0, \quad A = 1, \dots, K, \quad \alpha = 1, \dots, m \quad (84)$$

$$\varphi_{\alpha\beta} = 0, \quad \alpha, \beta = 1, \dots, m \quad (85)$$

The matrix C_{ij}^{-1} has the following property:

$$\sum_{j=1}^N C_{ij}^{-1} \tilde{\Omega}_{j\tilde{C}} = \sum_{\tilde{A}, \tilde{B} \in H} \frac{\tilde{\Omega}_{i\tilde{A}}}{\xi_i^2} \tilde{Q}_{\tilde{A}\tilde{B}}^{-1} \varphi_{\tilde{B}\tilde{C}} \quad (86)$$

which (together with (84) and (85)) in turn implies that

$$\sum_{j=1}^N C_{ij}^{-1} G_{j\alpha} \equiv 0, \quad \alpha = 1, \dots, m \quad (87)$$

This results in a solution (74) satisfying the linear constraints (80). In practice, to minimize noise in the factor model covariance matrix, the factor loadings Ω_{iA} should be chosen orthogonal to the matrix $G_{i\alpha}$ [16]:

$$\sum_{i=1}^N \Omega_{iA} G_{i\alpha} \equiv 0, \quad A = 1, \dots, K, \quad \alpha = 1, \dots, m \quad (88)$$

This is not required for the above “padding” trick, which works irrespective of (88).³¹

Another consideration is that in practice one often needs to impose upper and lower bounds on w_i :

$$w_i^- \leq w_i \leq w_i^+ \quad (89)$$

See [16] for a practically-oriented discussion. Assuming $w_i^- < 0$ and $w_i^+ > 0$, we can readily incorporate such bounds using the algorithm given in [16] for which the source code is given in [17]. The bounds (89) are simply imposed in optimizing the Sharpe ratio with the “effective” expected returns \hat{E}_i on the r.h.s. of (74) (but no bounds are imposed in (72)).

³¹ Also, the “padding” is needed only in C_{ij} on the r.h.s. of (74), not in the definitions (72).

Finally, for our backtesting purposes below, here we discuss how to include trading costs. Including nonlinear impact complicates the problem and is unnecessary for our purposes here. However, we can include linear trading costs. Below we will consider purely intraday strategies where the positions are established just once at the open and are liquidated just once at the close of the same trading day. For the stock labeled by i , let the linear trading cost per *dollar* traded be τ_i . Then including such costs in the case of optimizing the Sharpe ratio with the expected returns E_i amounts to replacing the expected return for the portfolio (3) by³²

$$E = \sum_{i=1}^N [E_i w_i - \tau_i |w_i|] \quad (90)$$

A complete algorithm for including linear trading cost in mean-variance optimization is given in, e.g., [16]. However, for our purposes here the following simple “hack” suffices. We can define the effective return

$$E_i^{eff} = \text{sign}(E_i) \max(|E_i| - \tau_i, 0) \quad (91)$$

and simply set

$$E = \sum_{i=1}^N E_i^{eff} w_i \quad (92)$$

I.e., if the magnitude for the expected return for a given stock is less than the expected cost to be incurred, we set the expected return to zero, otherwise we reduce said magnitude by said cost. This way we can avoid a nontrivial iterative procedure (see [16]), albeit we emphasize that this solution is only an approximation to the optimal solution. However, here we are already employing other approximations, so this way of treating linear trading costs is well-justified.³³

So, what should we use as τ_i in (91)? The model of [2] is reasonable for our purposes here. Let H_i be the *dollar* amount traded for the stock labeled by i . Then for the linear trading costs we have

$$T_i = \zeta \sigma_i \frac{|H_i|}{A_i} \quad (93)$$

where σ_i is the historical volatility, A_i is the average daily dollar volume (ADDV), and ζ is an overall normalization constant we need to fix. However, above we work

³² This is the expected return of the portfolio once it is established. In computing the P&L, we must take into account not only the establishing costs, but also the liquidating costs (so the total costs subtracted from the P&L are approximately double the establishing costs).

³³ For “multiply-optimized” strategies in (74), it may make sense to use more sophisticated approximations. For the sake of simplicity and not to overcomplicate things, we will use (91) here.

with weights w_i , not traded dollar amounts H_i . In our case of a purely intraday trading strategy discussed above, they are related simply via $H_i = I w_i$, where I is the total investment level (i.e., the total absolute dollar holdings of the portfolio after establishing it). Therefore, we have (note that $T_i = \tau_i |H_i| = \tau_i I |w_i|$)

$$\tau_i = \zeta \frac{\sigma_i}{A_i} \quad (94)$$

We will fix the overall normalization ζ via the following heuristic. We will (conservatively) assume that the average linear trading cost per dollar traded is 10 bps (1 bps = 1 basis point = 1/100 of 1%),³⁴ i.e., $\text{mean}(\tau_i) = 10^{-3}$ and $\zeta = 10^{-3}/\text{mean}(\sigma_i/A_i)$.

3.2 Backtests

Here we discuss some backtests. We wish to see how “multiply-optimized” strategies (74) for dollar-neutral intraday models compare with optimizing the Sharpe ratio (i.e., a “singly-optimized” strategy). For this comparison, we run our backtests as in [17]. For our C_{ij} (in all cases) we use heterotic risk models of [17]. The historical data we use in our backtests here is the same as in [17] and is described in detail in Subsections 6.2 and 6.3 thereof. The trading universe selection is described in Subsection 6.2 of [17]. We assume that the portfolio is established at the open with fills at the open prices; and ii) it is liquidated at the close on the same day – so this is a purely intraday strategy – with fills at the close prices. We include the transaction costs as discussed in Subsection 3.1 hereof. Furthermore, we include strict trading bounds (which in this case are the same as position bounds)

$$|H_i| \leq 0.01 A_i \quad (95)$$

We further impose strict dollar-neutrality on the portfolio, so that

$$\sum_{i=1}^N H_i = 0 \quad (96)$$

The total investment level in our backtests here is $I = \$20\text{M}$ (i.e., \$10M long and \$10M short), same as in [17]. For the Sharpe ratio optimization with bounds we use the R function `bopt.calc.opt()` in Appendix C of [17]. We use $\hat{b} = 1$ (see above) in “multiply-optimized” strategies. The backtest results are summarized in Table 1, which shows that the $n_{opt} = 2$ strategy outperforms the $n_{opt} = 1$ strategy (which is simply optimizing the Sharpe ratio). However, for higher n_{opt} it appears that we get – quite literally – “diminishing returns”.

³⁴ This amounts to assuming that, to establish an equally-weighted portfolio, it costs 10 bps.

4 Concluding Remarks

For long-only portfolios optimizing the Fano ratio effectively amounts to shifting the expected returns by a positive amount via (35), which results in fewer stocks being excluded from the portfolio due to the bounds (13) (including some stocks with negative expected returns E_i , for which effective returns \tilde{E}_i can be positive), i.e., in a more diversified portfolio³⁵ (compared with optimizing the Sharpe ratio).

However, for long-short portfolios this is a non-issue to begin with: the weights need not be nonnegative. As we discuss above, optimizing the Fano ratio in this case would be suboptimal. However, the Fano ratio optimization inspires considering modifications of optimizing the Sharpe ratio, such as (66) and (67). In this regard, the following comment is in order. Linearizing the hyperbolic tangent in (66) amounts to completely removing the sign “flip-flopping” issue discussed in Section 3, which (to a lesser degree) is present even when we replace the sign in (66) by the hyperbolic tangent in (67). The further reduction via (74) essentially amounts to simply combining multiple different alphas – even though here alphas are of a specific (“multiply-optimized”) form. However, more generally, combining multiple (even a large number of) different alphas yields higher returns and Sharpe ratios and lower turnover and higher cents-per-share (see, e.g., [22]).

Finally, let us mention that the Fano ratio arises in the context of statistical industry classifications via clustering techniques [21]. One question for clustering in the context of quant trading is what to cluster? Clustering returns is suboptimal. Naively, clustering normalized returns E_i/σ_i appears to be reasonable. However, as was argued and supported via backtests in [21], clustering E_i/σ_i^2 – i.e., the corresponding Fano ratios – is the optimal choice. Thus, clustering E_i/σ_i groups together stocks that are (to varying degrees) highly correlated in-sample. However, there is no guarantee that they will remain as highly correlated out-of-sample. Intuitively, it is evident that higher volatility stocks are more likely to get uncorrelated with their respective clusters. This is essentially why suppressing by another factor of σ_i in the Fano ratio E_i/σ_i^2 (as compared with E_i/σ_i) leads to better performance: inter alia, it suppresses contributions of those volatile stocks into the corresponding cluster centers [21].

³⁵ And also less skewed portfolio.

References

- [1] Adcock, J.C. and Meade, N. (1994) A simple algorithm to incorporate transactions costs in quadratic optimization. *European Journal of Operational Research* 79(1): 85-94.
- [2] Almgren, R., Thum, C., Hauptmann, E. and Li, H. (2005) Equity market impact. *Risk Magazine* 18(7): 57-62.
- [3] Bai, Z.D., Hui, Y.C. and Wong, W.K. (2015) Internet Bubble Examination with Mean-Variance Ratio. In: Lee, C.F. and Lee, J.C. (eds.) *Handbook of Financial Econometrics and Statistics*. New York, NY: Springer, pp. 1451-1465.
- [4] Bai, Z.D., Wang, K.Y. and Wong, W.K. (2011) The mean-variance ratio test – A complement to the coefficient of variation test and the Sharpe ratio test. *Statistics & Probability Letters* 81(8): 1078-1085.
- [5] Best, M.J. and Hlouskova, J. (2003) Portfolio selection and transactions costs. *Computational Optimization and Applications* 24(1): 95-116.
- [6] Bouchaud, J.-P. and Potters, M. (2011) Financial applications of random matrix theory: a short review. In: Akemann, G., Baik, J. and Di Francesco, P. (eds.) *The Oxford Handbook of Random Matrix Theory*. Oxford, United Kingdom: Oxford University Press.
- [7] Bowler, B. and Wentz, P. (2005) Portfolio Optimization: MAD vs. Markowitz. *Undergraduate Math Journal (Rose-Hulman Institute of Technology)* 6(2): 3.
- [8] Braga M.D. (2016) Alternative Approaches to Traditional Mean-Variance Optimisation. In: Basile, I. and Ferrari, P. (eds) *Asset Management and Institutional Investors*. Cham, Switzerland: Springer, pp. 203-213.
- [9] Cadenillas, A. and Pliska, S.R. (1999) Optimal trading of a security when there are taxes and transaction costs. *Finance and Stochastics* 3(2): 137-165.
- [10] Delbos, F. and Gilbert, J.C. (2005) Global linear convergence of an augmented Lagrangian algorithm for solving convex quadratic optimization problems. *Journal of Convex Analysis* 12(1): 45-69.

- [11] den Hertog, D. (1992) *Interior Point Approach to Linear, Quadratic and Convex Programming*. Mathematics and its Applications 277. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- [12] Fano, U. (1947) Ionization Yield of Radiations. II. The Fluctuations of the Number of Ions. *Physical Review* 72(1): 26.
- [13] Grinold, R.C. and Kahn, R.N. (2000) *Active Portfolio Management*. New York, NY: McGraw-Hill.
- [14] Janeček, K. and Shreve, S. (2004) Asymptotic analysis for optimal investment and consumption with transaction costs. *Finance and Stochastics* 8(2): 181-206.
- [15] Jansen, B. (1997) *Interior Point Techniques in Optimization – Complementarity, Sensitivity and Algorithms*. Applied Optimization 6. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- [16] Kakushadze, Z. (2015a) Mean-Reversion and Optimization. *Journal of Asset Management* 16(1): 14-40.
Available online: <http://ssrn.com/abstract=2478345>.
- [17] Kakushadze, Z. (2015b) Heterotic Risk Models. *Wilmott Magazine* 2015(80): 40-55. Available online: <http://ssrn.com/abstract=2600798>.
- [18] Kakushadze, Z. (2016) 101 Formulaic Alphas. *Wilmott Magazine* 2016(84): 72-80. Available online: <http://ssrn.com/abstract=2701346>.
- [19] Kakushadze, Z. (2017) On Origins of Bubbles. *Journal of Risk & Control* 4(1): 1-30. Available online: <http://ssrn.com/abstract=2830773>.
- [20] Kakushadze, Z. and Yu, W. (2016a) Multifactor Risk Models and Heterotic CAPM. *The Journal of Investment Strategies* 5(4): 1-49.
Available online: <http://ssrn.com/abstract=2722093>.
- [21] Kakushadze, Z. and Yu, W. (2016b) Statistical Industry Classification. *Journal of Risk & Control* 3(1) (2016) 17-65.
Available online: <http://ssrn.com/abstract=2802753>.
- [22] Kakushadze, Z. and Yu, W. (2017a) How to Combine a Billion Alphas. *Journal of Asset Management* 18(1): 64-80.
Available online: <http://ssrn.com/abstract=2739219>.

- [23] Kakushadze, Z. and Yu, W. (2017b) Statistical Risk Models. *The Journal of Investment Strategies* 6(2): 1-40.
Available online: <http://ssrn.com/abstract=2732453>.
- [24] Kellerer, H., Mansini, R. and Speranza, M.G. (2000) Selecting Portfolios with Fixed Costs and Minimum Transaction Lots. *Annals of Operations Research*, 99(1-4): 287-304.
- [25] Konno, H. and Yamazaki, H. (1991) Mean-Absolute Deviation Portfolio Optimization Model and Its Applications to Tokyo Stock Market. *Management Science* 37(5): 519-531.
- [26] Lo, A.W. and Patel, P.N. (2008) 130/30: The New Long-Only. *The Journal of Portfolio Management* 34(2): 12-38.
- [27] Lobo, M.S., Fazel, M. and Boyd, S. (2007) Portfolio optimization with linear and fixed transaction costs. *Annals of Operations Research* 152(1): 341-365.
- [28] Markowitz, H. (1952) Portfolio selection. *The Journal of Finance* 7(1): 77-91.
- [29] Michaud, R. and Michaud, R. (2008) Efficient Asset Management: A practical Guide to Stock Portfolio Optimization and Asset Allocation. (2nd ed.) Oxford, UK: Oxford University Press.
- [30] Mokkavesa, S. and Atkinson, C. (2002) Perturbation solution of optimal portfolio theory with transaction costs for any utility function. *IMA Journal of Management Mathematics* 13(2): 131-151.
- [31] Murty, K.G. (1988) *Linear Complementarity, Linear and Non-linear Programming*. Sigma Series in Applied Mathematics 3. Berlin: Heldermann Verlag.
- [32] Pang, J.-S. (1983) Methods for quadratic programming: A survey. *Computers & Chemical Engineering* 7(5): 583-594.
- [33] Patel, N.R. and Subrahmanyam, M.G. (1982). A Simple Algorithm for Optimal Portfolio Selection with Fixed Transaction Costs. *Management Science* 28(3): 303-314.
- [34] Rockafellar, R.T. and Uryasev, S. (2000) Optimization of Conditional Value-At-Risk. *The Journal of Risk* 2(3): 21-41.

- [35] Roy, O. and Vetterli, M. (2007) The effective rank: A measure of effective dimensionality. In: *Proceedings – EUSIPCO 2007, 15th European Signal Processing Conference*. Poznań, Poland (September 3-7), pp. 606-610.
- [36] Sharpe, W.F. (1966) Mutual Fund Performance. *Journal of Business* 39(1): 119-138.
- [37] Sharpe, W.F. (1994) The Sharpe Ratio. *The Journal of Portfolio Management* 21(1): 49-58.
- [38] Shreve, S. and Soner, H.M. (1994) Optimal investment and consumption with transaction costs. *The Annals of Applied Probability* 4(3): 609-692.

Table 1: Simulation results for the “multiply-optimized” strategies (defined in Eq. (74) in Section 3) with bounds and costs. ROC = Return-on-Capital, SR = annualized Sharpe Ratio, CPS = cents-per-share. Note that $n_{opt} = 1$ corresponds to vanilla optimization of the Sharpe ratio (i.e., a “singly-optimized” strategy).

n_{opt}	ROC	SR	CPS
1	35.37%	13.65	1.74
2	36.62%	15.43	2.02
3	34.00%	15.39	2.06
4	26.38%	11.80	1.69
5	17.00%	7.27	1.09

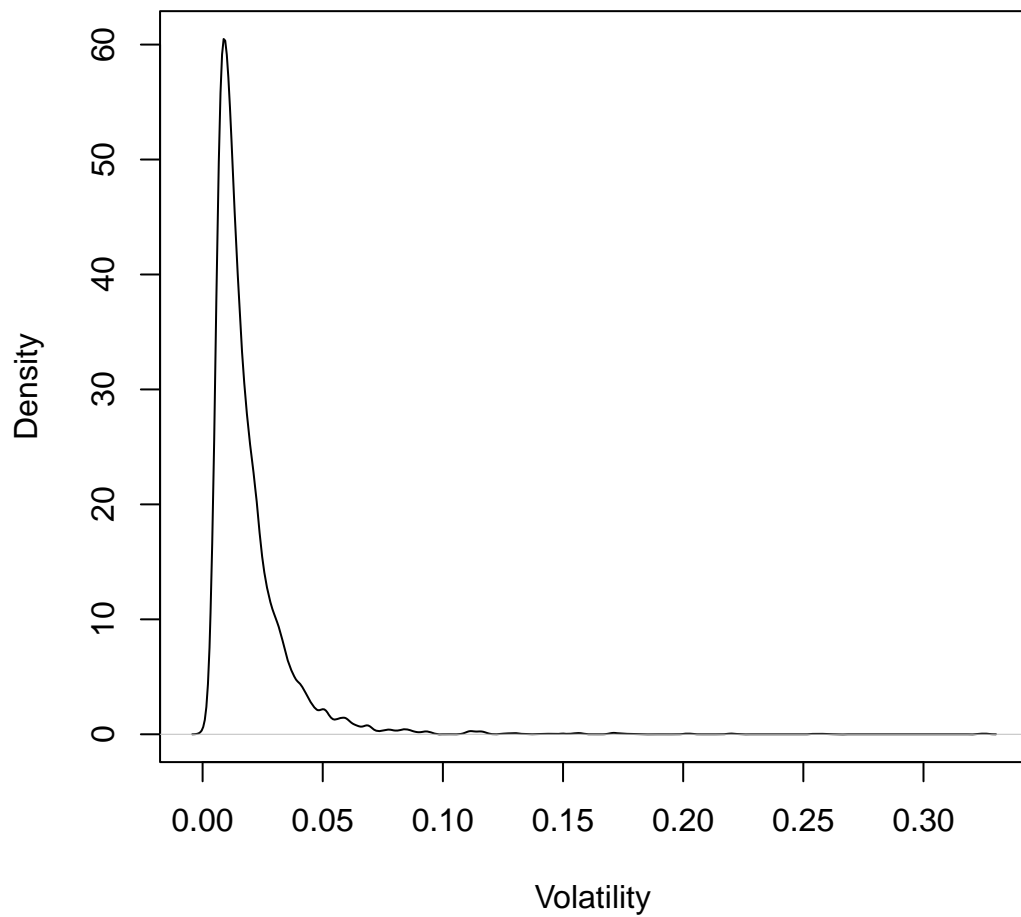


Figure 1: Density of volatility for the universe of 3810 stocks (see Section 2.6).

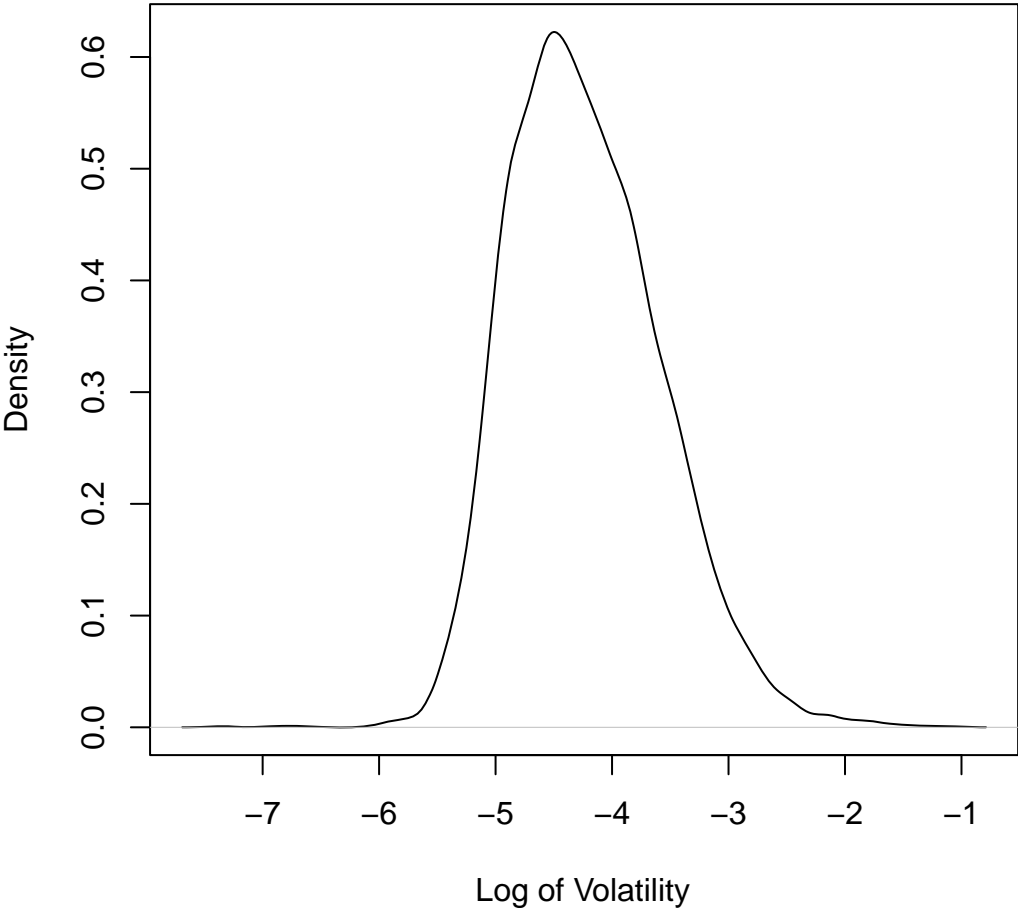


Figure 2: Density of log-volatility for the universe of 3810 stocks (see Section 2.6).

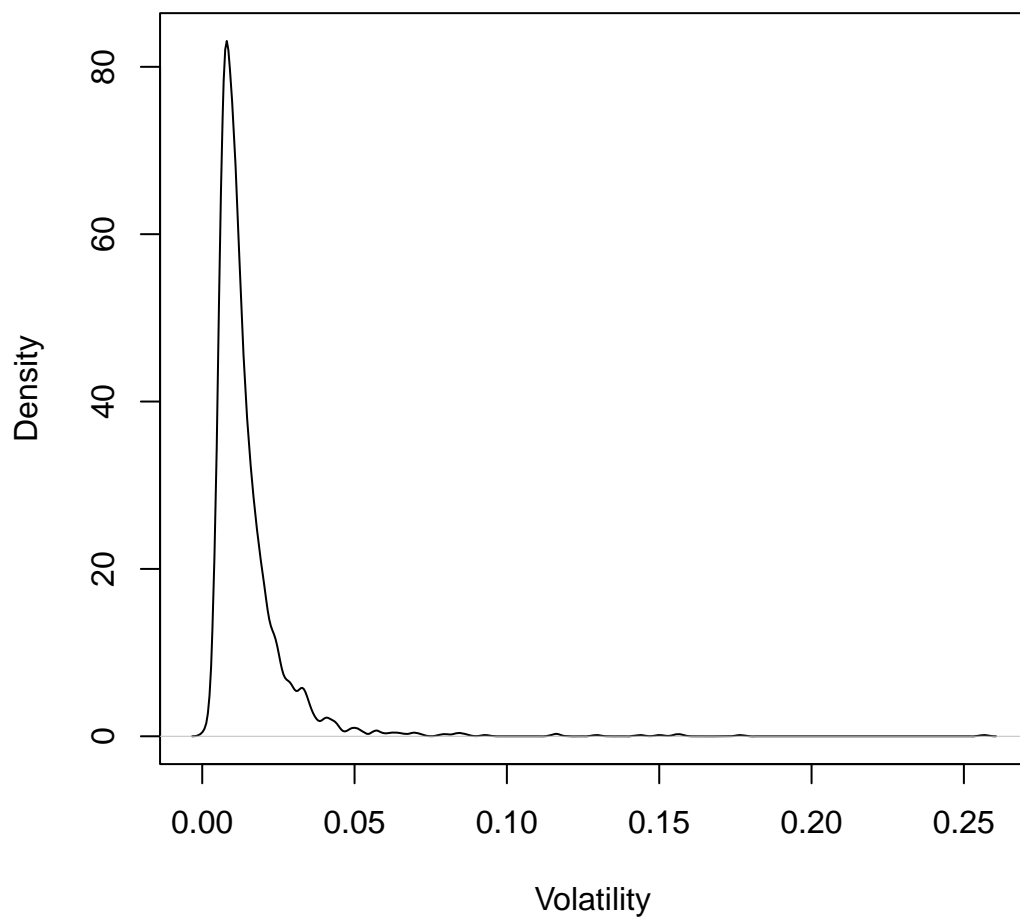


Figure 3: Density of volatility for the universe of 2000 most liquid stocks (see Section 2.6).

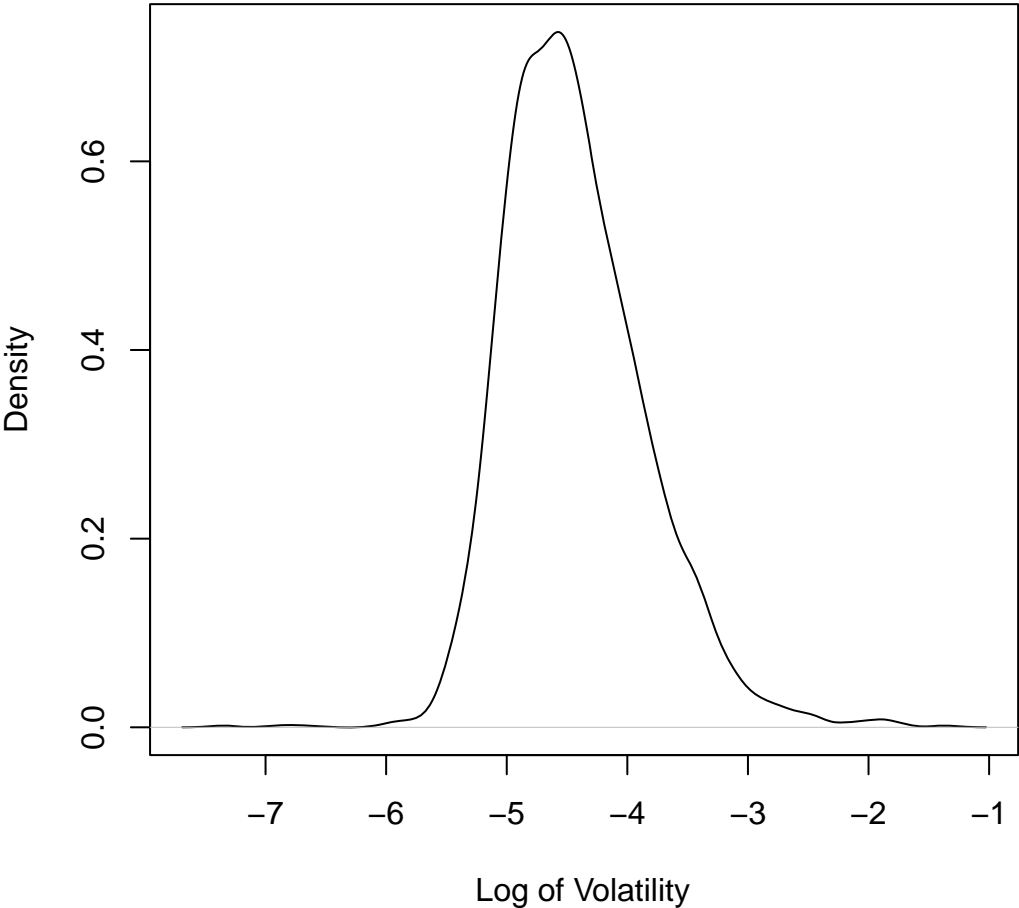


Figure 4: Density of log-volatility for the universe of 2000 most liquid stocks (see Section 2.6).